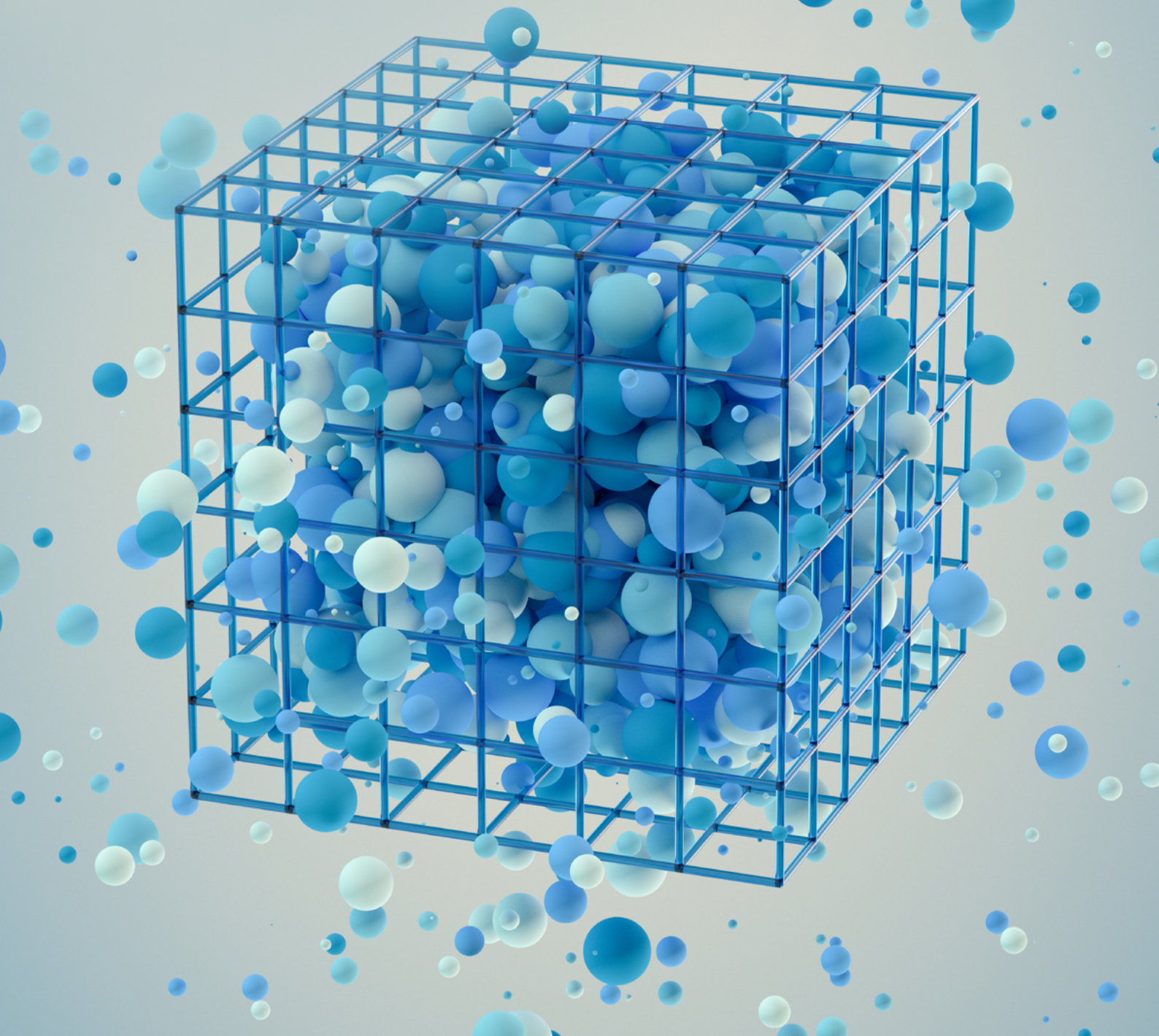
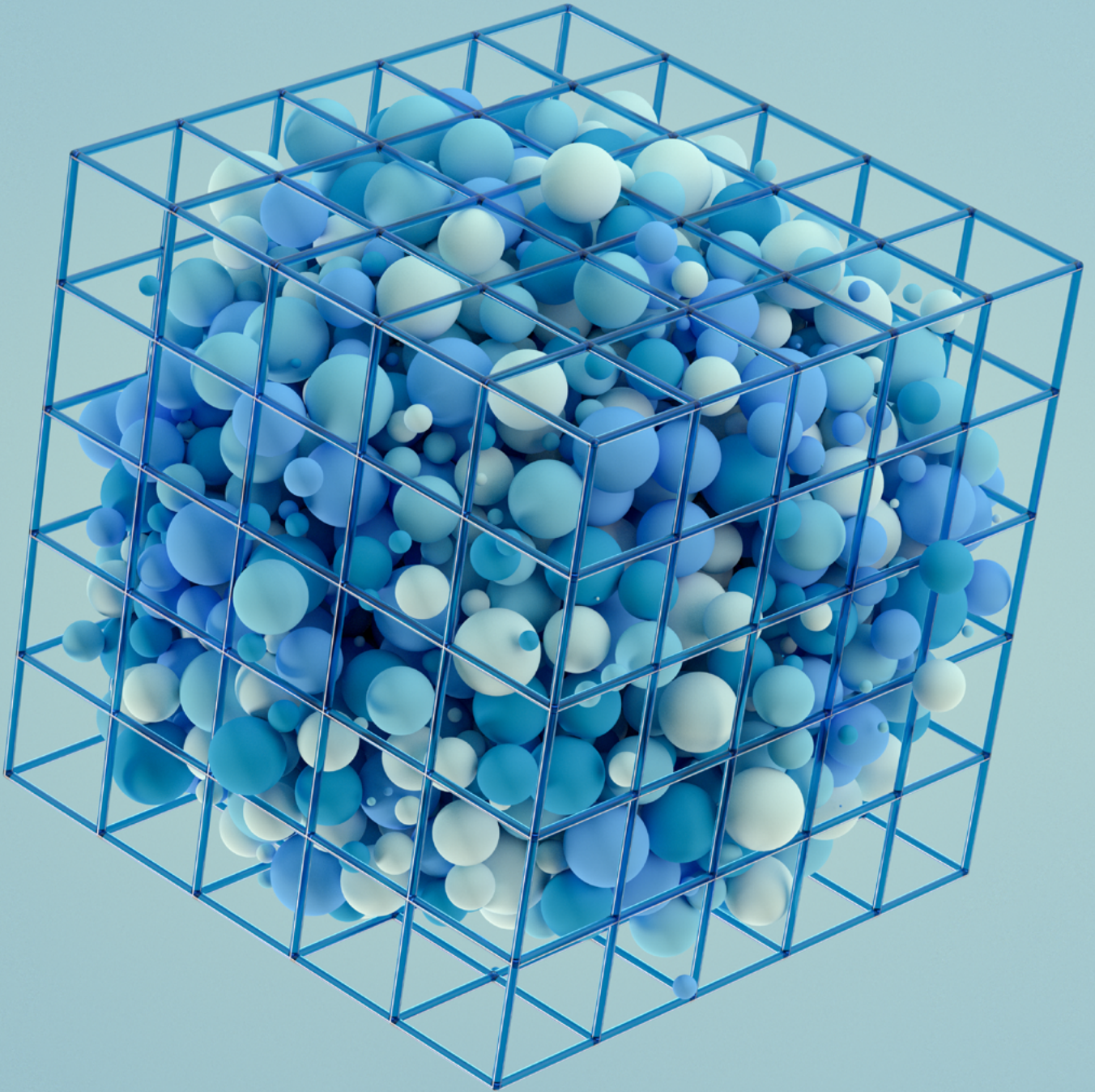


# Moving to and within the cloud

The importance of using data lineage and blueprints to avoid unintended consequences





# | Contents

## **03 Introduction**

**04** Overview

**04** A further note on the CDMC framework

**04** When you move to the cloud: a regulatory perspective

**05** General considerations to take before moving to the cloud

## **05 Data lineage**

## **07 Using blueprints for scenario planning, impact analysis and more**

**07** Blueprints: an overview

**07** Reducing risk with robust impact analysis

**08** Essential views on your cloud transformation program: 'as-is' and 'to-be' lineage

**09** Data privacy, data sharing and other regulatory responsibilities

**09** Business continuity

**09** Business opportunities

## **10 A non-exhaustive checklist**

## **11 Conclusion**

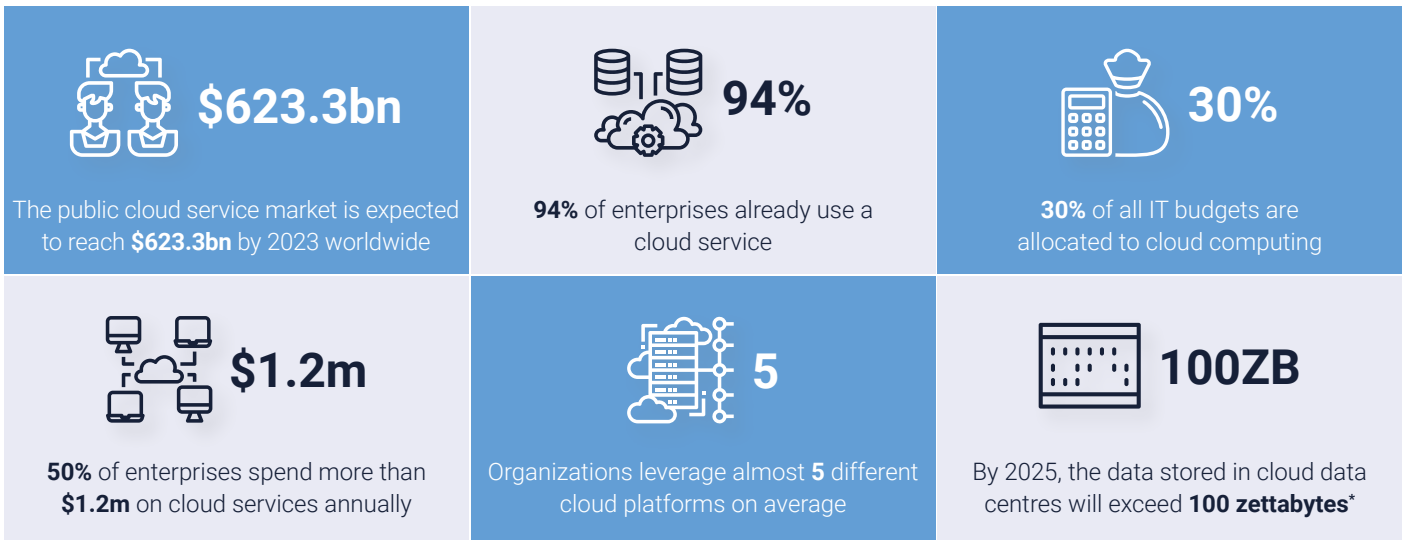
## **12 About Solidatus**

# Introduction

## Overview

Take a moment to reflect on the past five years of technological innovation. During that time, it would be hard to find a better or more universal example of a case of ‘when, not if’ than the issue of moving your business systems to the cloud.

These are just a handful of statistics from a recent Web Tribunal [article](#):



\* (A zettabyte is 10<sup>21</sup> bytes. Or, to put it in lay terms, a LOT.)  
(1ZB = 1bn terabytes)

So, it's highly likely that you're already some way there. Few businesses haven't started the process. Many operate a hybrid model with their tech straddling the cloud and on-prem servers. But that's the thing: it's a process, not an event, and a highly complex, risky and time-consuming one at that.

You might even call it a discipline – almost in the monastic sense of the word – which, once adopted, demands a continuing daily obligation in terms of observance and behaviour.

The overriding objective of data management in its broadest sense is optimal support for the business. So, moving data or systems to the cloud is not a one-off decision but something requiring continual monitoring to ensure that yesterday's decision is still right for the business today. To do that in a way that preserves your sanity, you need to understand the entirety of your data estate and how it maps to your business functions.

There's also no real end date; even when the bulk of your operations are in the cloud, which cloud is it? There are several to choose from, all with their strengths and weaknesses, so there's a good chance you'll choose more than one. Which of your systems should go where? How will these systems interact with each other once they're up in the heavens, and what changes might you have to make down the line?

This is before you look at the ESG imperatives in relation to the environmental footprint of cloud data centres and blockchain, which is perhaps a story for another day.

None of these questions is easy to answer. And none can be treated in isolation. But they must all be addressed. So, the overarching question isn't whether but *how*.

There are literally thousands of free online resources on the myriad considerations of cloud migration. They address a wide range of issues, and we'd of course recommend you read some of them.

But in this document, we focus on the under-discussed but highly valuable concept of lineage – for your data, systems and processes – and, by extension, the idea of data blueprints, all of which ties in with the EDM Council's CDMC (Cloud Data Management Capabilities) framework, which we wrote about on our [blog](#).

That ongoing initiative comprised six major categories:

- 1 Governance and accountability
- 2 Cataloguing
- 3 Accessibility
- 4 Protection
- 5 Data lifecycle
- 6 Data and technical architecture

And from it emerged five objectives for data lineage:

- 1 To implement automated functionality that identifies processes that move data.
- 2 To record data lineage metadata for data movement processes that are discovered automatically.
- 3 To ensure lineage auto-discovery identifies processes that move data across jurisdictions, availability zones and physical boundaries.
- 4 To ensure lineage auto-discovery is enabled in hybrid and multiple cloud environments, and identify data movement between those environments.
- 5 To define and implement processes for the review of auto-discovered lineage information.

As you read this whitepaper, you'll see how some of these themes dovetail with what we go on to discuss, but there's much else to consider, as you'll discover...

## A further note on the CDMC framework

Our [blog post](#) provides a good overview of the EDM Council's lineage recommendations in relation to cloud migration. But if you have time, it's worth reading about their full Cloud Data Management Capabilities Framework, a document on which you can download on the EDM Council [website](#).

In particular, Section 6.2 on understanding data provenance and lineage, which runs for several pages from page 143, provides invaluable advice for data practitioners, complementing our white paper.

Its 'scoring' tables in the same section should also be incorporated into your programme of activities.

## When you move to the cloud: a regulatory perspective

Before we proceed, it's worth pausing for a moment to ask what data and systems we always retain, however it is that we move them to the cloud. We're not selling the concept, and the technical benefits that serve a dispersing workforce – reliability, scalability, flexibility, consistent processing – are widely understood. But from data management and governance perspectives, the simple truth is that most regulators are now in the cloud.

And these regulators are increasingly unwilling to accept infractions in the cloud as the responsibility of the cloud provider, rather than the client using cloud services. That being so, only a complete understanding of how your cloud usage is arranged can provide an adequate defence against inadvertent transgression.

## General considerations to take before moving to the cloud

Before embarking on any cloud migration journey, you must:

- Understand the key data and systems, and therefore the greatest migration risks for your organization.
- Look at data sharing and permissions.
- Reflect on change management throughout your migration.
- Document your data estate throughout the process.
- Demonstrate through lineage documentation and reporting the control of the relationships with third parties and outsourcing arrangements.
- Reduce operational risk from extensive manual efforts related to data sourcing, standardization and analysis and report production.
- Reduce technology risk and increased operating expenses or project costs due to redundant sources of data and data silos.
- Improve operational reporting and real-time access to high-quality data.

We'll explore some of these in more depth as we ask the fundamental question: what does 'good' look like?

The simple answer is 'good' is a complete and genuine understanding of: where your data came from; its journey; where it is now; any transformations it's undergone or will undergo; and where it's going. It all starts with lineage.

## | Data lineage

A central theme of this document is lineage. It weaves its way into every aspect of data and systems management, and, by extension, cloud migration.

The concept of data lineage is very straightforward: it's simply a description of a piece of data's journey, one that outlines where it first came into your organization's databases and systems, and its stops along the way.

But it gets complicated for several reasons, including:

- Changes it undergoes along the way – Mike Turner on your company's Slack account is Michael Turner on the payroll
- Determining whether similar records in different databases relate to the same entity – Mike and Michael *are* one and the same, but B. Andrew Brown and Andrew Brown could well be different
- Split paths that a piece of data takes – linear paths are rarely smooth and one-dimensional
- Version control and the 'temporal' nature of lineage

And this is before you look at its many possible onwards journeys, some pre-determined, others up for debate.

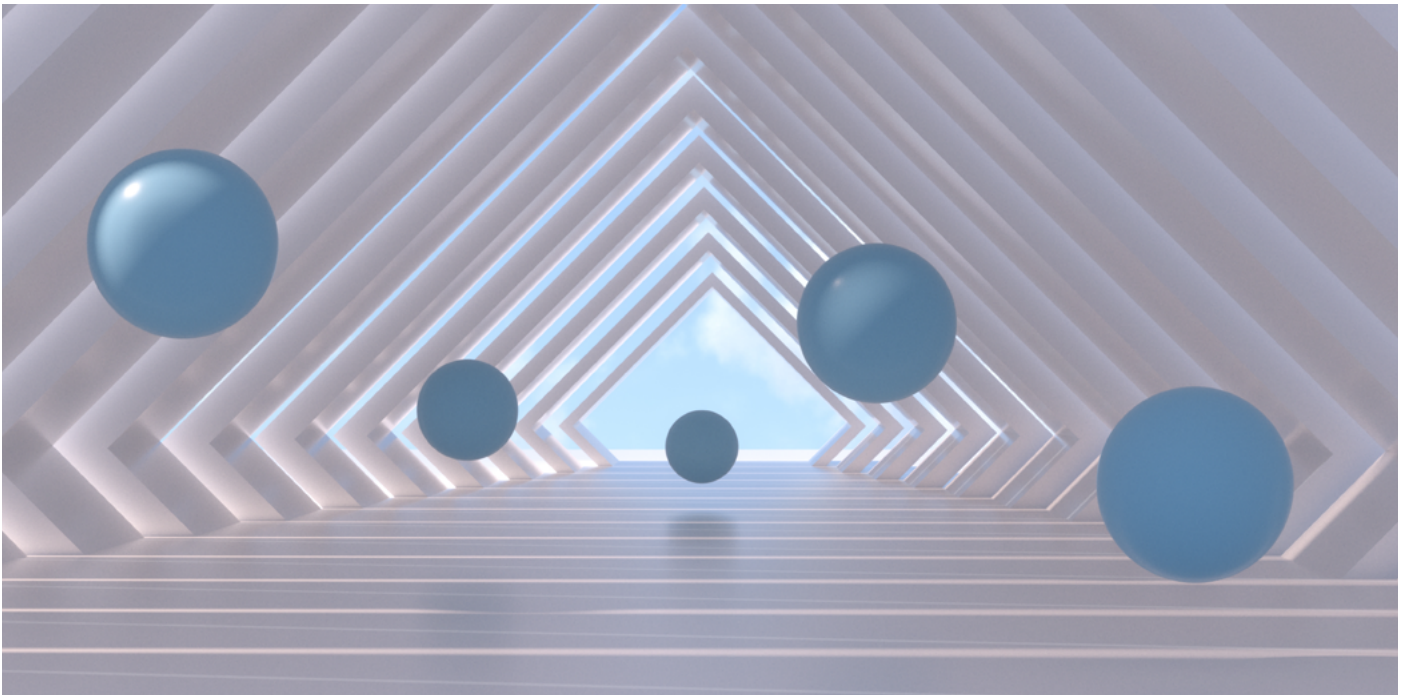
The larger the organization, and the more complex and diverse its systems, the greater the data lineage headache. When the cloud is brought into the mix, along with the local data regulations associated with wherever your cloud is hosted, you can safely conclude that the picture isn't simplified.

The only way to keep your head is to model your lineage with a software solution that allows you to visualize it and, by applying rules that you set, isolate specific paths that cut out the noise.

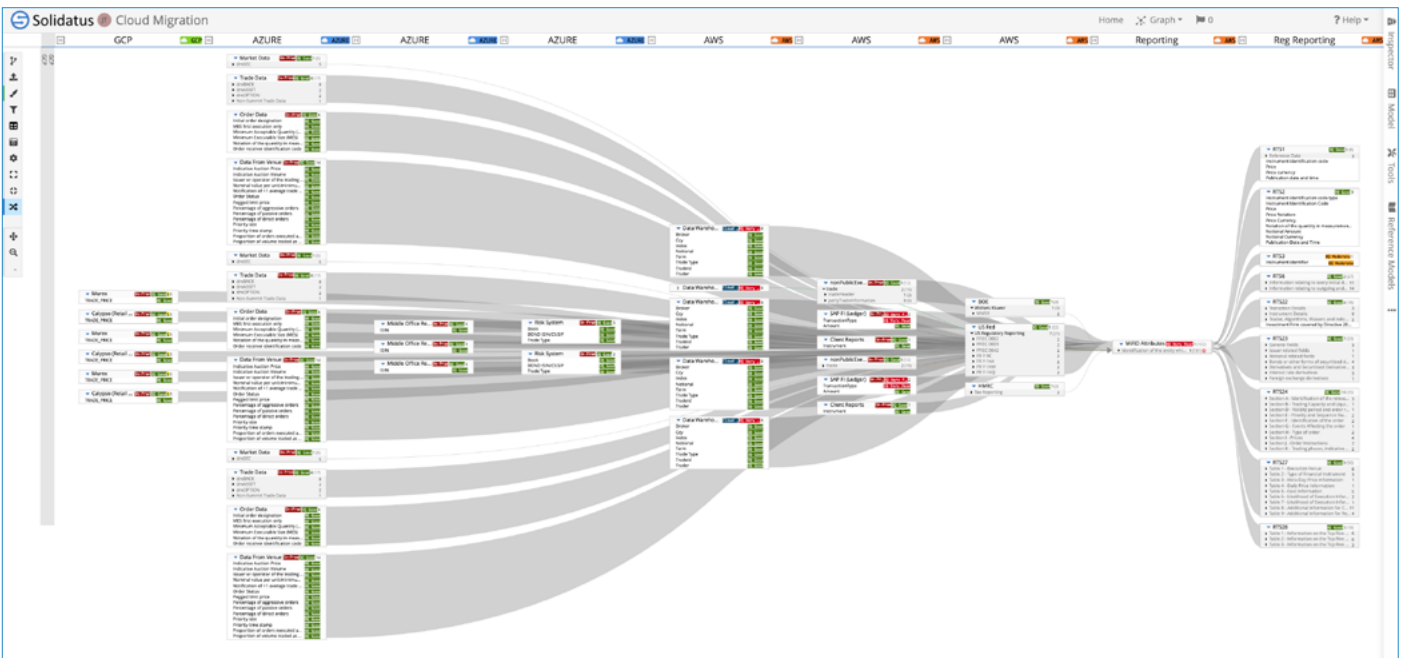
Put simply, a lineage model – such as those you can build in Solidatus ([solidatus.com](https://solidatus.com)) – is a visualization that shows you the interplay of data (or other types of 'objects' that you map) and the uses to which it is put within a context framed by the modeler.

It's also worth flagging up a related type of model, one that Solidatus also supports: the reference model. A reference model provides a hierarchical representation of concepts that can be cross-referenced with entities in lineage models or other reference models. A lineage model can be linked to one or more reference models describing business terminology, regulatory principles, and more.

This cross-referencing is crucial when considering the jurisdiction point we raise in the overview above.



Before we move on to the next section, which builds on these ideas, take a look at our [cloud data management page](#), where you can zoom in on the sample lineage model reproduced here.



# Using blueprints for scenario planning, impact analysis and more

If you're clear on the principles of data lineage, then scenario planning and impact analysis easily follow as extensions to it. These twin disciplines can be aided by blueprints. But what do we mean by that term?

## Blueprints: an overview

What is a blueprint?

In the world of construction, a blueprint is a two-dimensional drawing that provides a visual representation of a building's layout, dimensions, component placement, electrical wiring and construction materials. Drawn up by architects or engineers, blueprints allow you to quickly check and identify different building elements and verify compliance with building codes.

A blueprint follows a carefully thought-out, logical sequence of steps. Drawing pages in blueprint sets are arranged in a predictable fashion, and blueprint symbols and lines have highly specific meanings.

Imagine if instead of a building inspection, you were required to undergo a business inspection.

Could you visually lay out the dimensions and components that inform your business decisions and drive your operations?

Are you confident that you have structural integrity, and that your data governance and management capabilities provide the appropriate level of safety and soundness required to verify compliance with regulatory obligations, reporting requirements and privacy standards?

Traditional lineage approaches can be daunting for any organization. Having to source and then document the sources of information that describe the processes, people, systems, data, reports and controls seems unattainable. Especially considering that everything that *needs* to be documented has likely never *been* documented.

With the right lineage software solution, you can ingest metadata across key systems, providing a view of the federated collection and documentation of key processes that drive your business.

Put simply, a blueprint – in the context of data management generally and cloud migration specifically – is an interactive, living visualization of how your data flows and its connection to the obligations that regulate it, your policies that guide it and your processes that create or use it – both now and at other points in time.

## Reducing risk with robust impact analysis

The value of a good blueprint extends far beyond regulatory reporting and controls, though we'll touch on these later in this chapter. The business benefits associated with the ability to conduct a thorough impact analysis as part of major transformation programs are significant.

To maximise efficiency and avoid unintended consequences, this is hugely applicable to cloud migrations and the phased approach that you're likely to take – or are already undertaking.

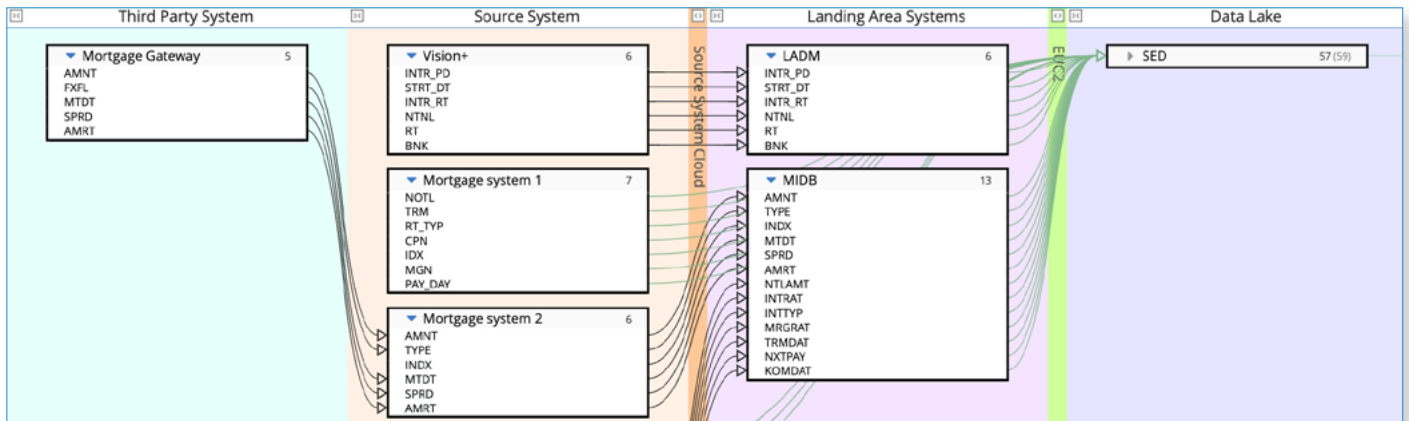
With access to the right software, you can see the magnitude and scope of downstream impacts of moving to or within the cloud – and it's of course much cheaper to assess these multiple, bitemporal 'what if' scenarios *before* they potentially impact your business than after.

The insights and intelligence from having this landscape view, and in understanding the level of interconnectedness of your applications and data repositories, will enable you to budget appropriately and plan accordingly – accelerating your program timelines by making the 'unknown known' much more quickly than traditional current state assessments.

## Essential views on your cloud transformation program: 'as-is' and 'to-be' lineage

One of the advantages of having data lineage visualized is being able to see the impacts of changes to data architecture. Imagine that we work for a company that plans to migrate data housed in multiple on-prem systems to the cloud. Specifically, we're concerned with our credit card (Vision+) system and our four mortgage systems. Each feeds into a landing area, a data lake and a series of other systems, but in this case, we are only concerned with investigating the impact of migrating our on-prem source system to the cloud.

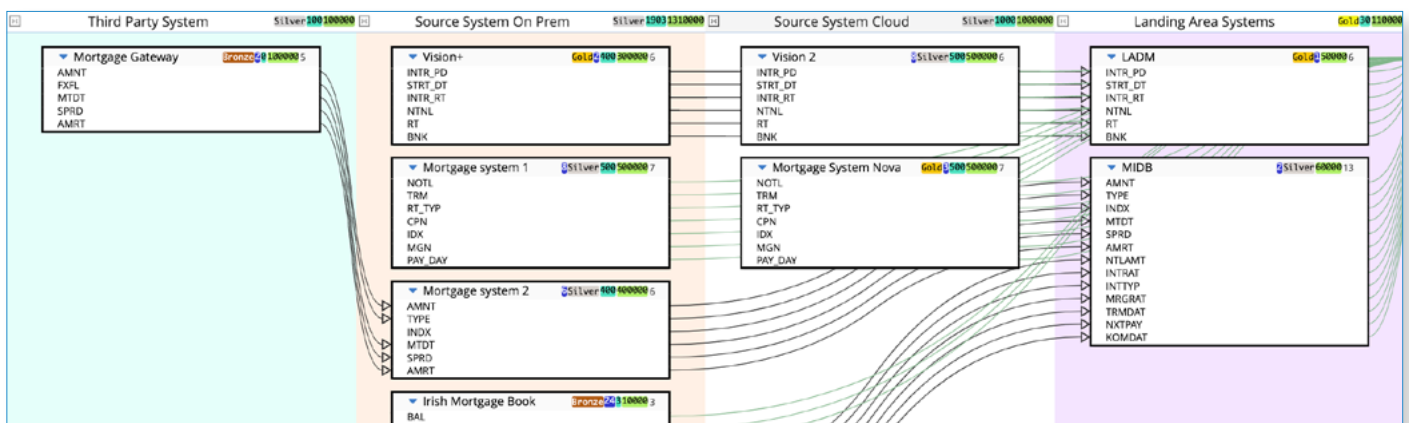
Several variables are of interest: cost, staff numbers, system ratings and service level agreements (SLAs). Below is a view of our 'as-is' data lineage.



Let's produce a view for our 'to-be' data lineage. By creating the layer titled: Source System Cloud, we can explore and compare our variables of interest against our current on-prem source system.

Using a tool like Solidatus, we can create a series of display rules that enables a user to illuminate parts of a model in order to highlight entities and expose additional metadata to enhance the user's view of that model.

Here's an example of a 'to-be' view:



You can read more about using display rules and variables to create different views on our website.

To reiterate, the crucial point is that it's far better to understand and then migrate than it is to migrate and then try to understand; decisions must be scientifically informed, not based on gut feel or, worse, ignorance.

We've looked at the *how* of impact analysis. Let's take a look at the *why*, or at least some of the key reasons.

## Data privacy, data sharing and other regulatory responsibilities

While this white paper doesn't focus or dwell on regulations, it goes without saying that you must include them in your plans when migrating to the cloud; and a lineage-based approach will help you keep track of what you need to consider vis-à-vis local regulations for wherever your cloud servers are hosted.

When it comes to reporting your activity, we should note that a regulator wouldn't ask for information that shouldn't already be known to the business, for business as opposed to regulatory purposes.

It all comes back to using blueprints for better decision-making and leveraging of expenditure to improve business data in pursuit of accurate regulatory submissions. If you get this right, your regulatory reporting will follow much more easily.

Which brings us on to the next two sections, which go hand-in-hand: business continuity and business opportunities.

## Business continuity

Businesses operate like flywheels. They must keep turning. To ensure business continuity, a data lineage and blueprint approach will – among a great many other things – help you:

- Establish mission-critical and non-mission-critical systems, helping you prioritise your road map.
- Minimize the number and frequency of migrations you have to do.
- Minimising downtime.
- Preserve the necessary elements of your legacy systems.
- Plan for disaster recovery; and
- Keep your cybersecurity systems in check.

But it's not just about reducing risk; it's about creating opportunities, something that a well-thought-through lineage-based approach to cloud migration will promote.

## Business opportunities

With a well-defined blueprint, you'll save time and money in your cloud migration planning. But this almost goes without saying. Let's take a very quick look at some of the other business opportunities of this approach.

- Reducing time and costs is dependent on increased efficiency – a lineage-based approach to cloud migration will eliminate redundancy with its attendant overheads.
- It will also reveal cross-selling opportunities, meaning you can optimize your client relationships.
- Enterprise-wide pattern appreciation is elevated, bringing with it opportunities in the artificial intelligence and machine learning space.
- From a business perspective, you'll almost certainly want to phase your migration to the cloud; this approach will help you optimize this phasing, as we touch on in the impact analysis section above.
- As there are some commercial implications for spreading across more than one cloud or more than one cloud provider, this must be factored into your planning, something that's aided by a lineage-first approach – see Cloud Data Management Capabilities Framework, as mentioned earlier.
- As we also said earlier, this approach facilitates more proactive data governance, which will reduce tail-chasing when the regulators come knocking.
- In its broadest application, you'll increase your data discovery, which unearths numerous secondary operational and business benefits.
- Perhaps above all – though at the end of this list! – this blueprint approach to cloud migration will help your organization foster a better data culture, which will permeate your whole business.

## A non-exhaustive checklist

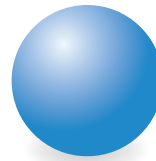
We've covered a lot of ground but before we conclude by reiterating some of the *benefits* of a lineage-first, blueprint-based approach to cloud migration, let's lay out a non-exhaustive checklist of the main points to consider or *steps to take* before diving into any cloud migration project with reference to a software solution that can help these activities.



Map your data, processes and systems in a lineage model or models.\*



Map the people you need to work on your migration project(s), not just the data and systems.



Set up rules within your model so that you can isolate work stream-related views.



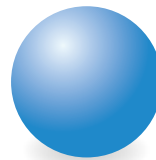
Add rich metadata so that you can create a fuller picture of your 'as-is' view and 'to-be' scenarios.



Create reference models, which can be cross-referenced with and linked to your lineage models, for a deeper understanding.



Review your work with your team through the visualisations that your models offer.



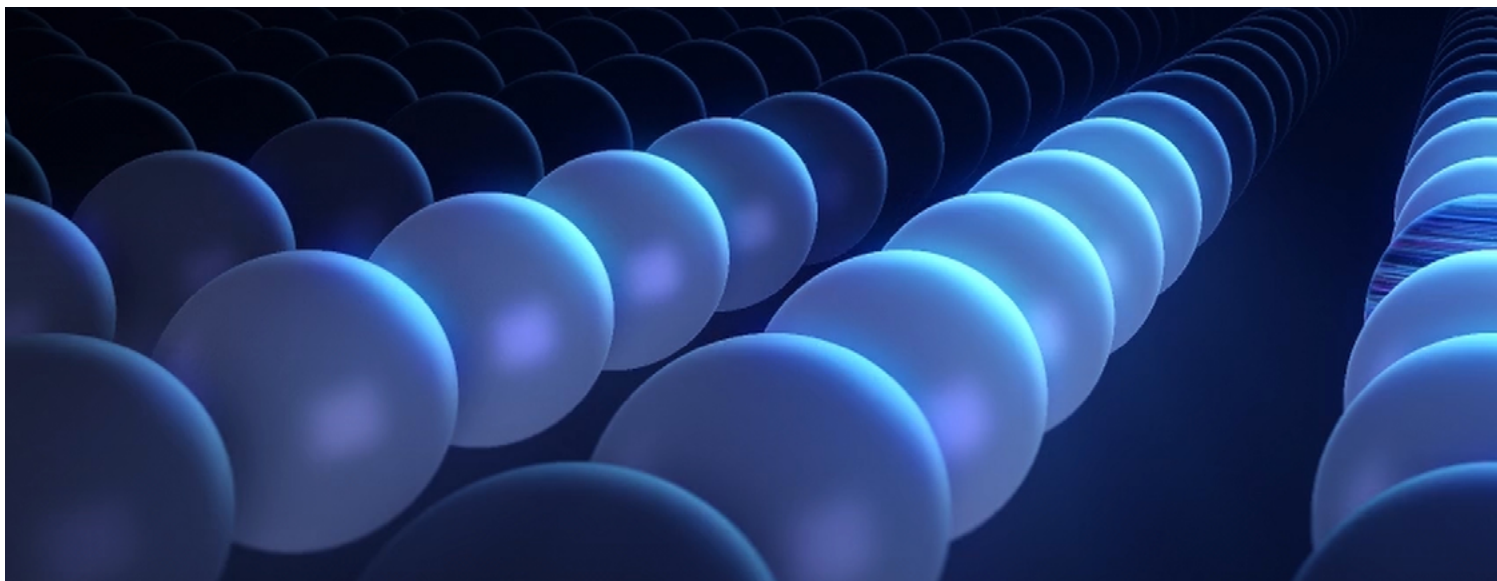
Explore these various scenarios as part of a thorough program of impact analysis.

You should also review the data lineage controls laid out on the last page of [Cloud Data Management Capabilities Framework](#), which include:

- **Risks addressed:** 'Data cannot be determined as having originated from an authoritative source resulting in a lack of trust of the data, inability to meet regulatory requirements, and inefficiencies in the organization's system architecture'.
- **Drivers and requirements:** 'Organizations need to trust data being used and confirm that it is being sourced in a controlled manner. Regulated organizations produce lineage information as evidence that the information on regulatory reports has been taken from an authoritative source for that type of data.'
- **Legacy and on-prem challenges:** 'Lineage information is produced manually by tracing the flow of data through systems from source to consumption. The cost of this approach and the consequences of producing incorrect data can be significant'. Using the right software can help reduce this manual pain.

Above all, foster a more collaborative data management culture. If you can articulate convincingly to your colleagues that data management is just a form of business management, it will be hard for them to refuse to collaborate. This means helping them clearly visualize business interdependencies.

\*Hopefully, you won't be starting from scratch, as these are used for efficient data management and governance well beyond the world of cloud migrations.



## Conclusion

By way of conclusion, let's simply consider the benefits that the right software can bring to your cloud-migration endeavours.

### **Building a complete cloud migration strategy**

With the right data lineage solution, you can experiment on a linked snapshot of the current view and assess the impact of making changes as regulations evolve, and as new data sources and organizations come online, preventing data duplication and wasted resources.

### **Protecting your reputation**

Mitigate the significant reputational risks associated with a data breach and avoid regulatory fines and sanctions by mapping data privacy, access and retention rules across all your business entities, jurisdictions, data categories, usages and systems – delivering major time savings and efficiency increases.

### **Maintaining a diverse cloud-services architecture**

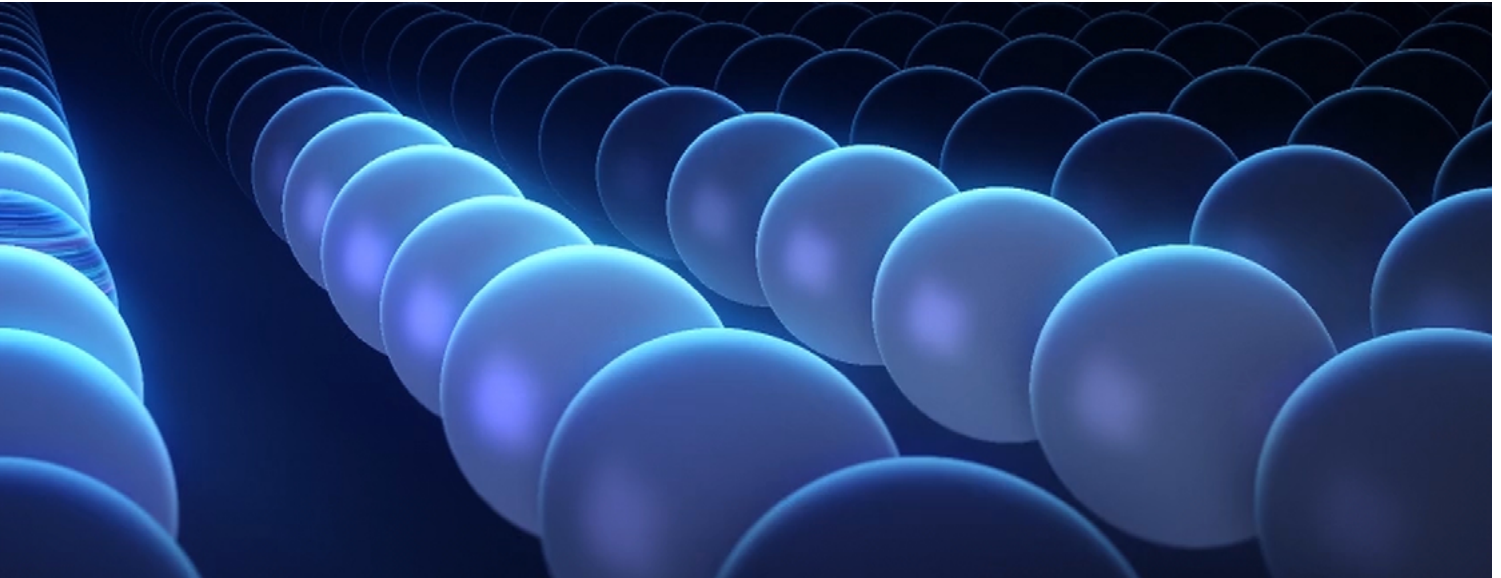
Lineage-first technology enables you to dynamically connect and visualize complex data relationships, break down silos, and avoid vendor lock-in and ensure compliance, which is particularly important if you're responsible for a global organization with multi-cloud data infrastructures.

### **Collaborative data management**

The right software allows data management to be distributed, federated and democratized. You can enable subject matter experts to discover, document, validate and disseminate information and expertise, while unlocking actionable insights behind data.

### **Transforming your organization**

Use initial investment in cloud transformation to deliver a 360° view across your organization to elevate and transform its data capabilities, improving data management and governance to increase revenue streams.



## | About Solidatus

### Overview

Solidatus is an innovative data management solution that empowers organizations to connect and visualize their data relationships, simplifying how they identify, access and understand them. With a sustainable data foundation in place, data-rich enterprises can meet regulatory requirements, drive digital transformation, capture business insights and make better, less risky and more informed data-driven decisions.

Solidatus' powerful metadata management technology is seen as a critical development in data management software – one that matches the complex needs of modern business. Launched in 2017, Solidatus is the chosen data management tool for both the regulators and the regulated. Its clients and investors include top-tier global financial services brands such as Citi and HSBC, healthcare and retail organizations as well as government institutions. For more information, visit [solidatus.com](https://solidatus.com).

### In relation to cloud migration, Solidatus enables you to:

- Capture both technical and business data lineage to foster an understanding based on a common language
- Conduct scenario planning and impact analysis to more accurately quantify impacts to your upstream and/or downstream processes, applications and data flow
- Connect the dots on disparate pieces of reporting and analysis, and formulate a cohesive understanding of your business performance drivers

## | Request a demo

We'd love to show you how Solidatus can benefit you.

Sign up for a demo on our website at [solidatus.com](https://solidatus.com) or email us on [hello@solidatus.com](mailto:hello@solidatus.com).

